Stretched exponential as one of the alternatives for the power law to fit ranked species abundance data

Gerben Straatsma, Simon Egli
Aquatic Ecology and Water Quality Management Group, Wageningen University and Research Center; Forest Dynamics, Swiss Federal Research Institute WSL

Abstract
A central object in community ecology is species abundance distribution. We are interested in the power law and its allies for ranked species abundance data. We collected 12 large data sets consisting of many samples. The preliminary fitting result makes a robust impression (12 systems at three scales of integration) that the stretched exponential is an interesting alternative for the power law. For further work, advanced statistics are required. Not only ‘our’ data but, quite often, other data as well consist of sample × species cross-tables. With cross-tables, also ‘within species over samples’ characteristics can be studied. An integrated view on data patterns in multi-sample sets may help to identify generative processes for and the formulation of a relatively simple model for species abundance data.

Objective
The direct aim is to generate interest again in the power law, especially in its allies like the stretched exponential, for species abundance data. Further reaching aims, to be tackled in the future and for which we introduce a frame, are (i) to generate the interest of community ecologists in the generative processes of the power law and allies (that have been studied in other fields of science) and (ii) to complete the quest for a simple yet meaningful equation/model for the SAD.

Introduction
A central object in community ecology is species abundance distribution (SAD). It has been studied for over a century since Raunkiaer (in [1]). The review of an expert group [2] is a benchmark for properties and generative theory. The edited volume [3] on biological diversity is, among others, an update. For applications of the SAD, like the measurement of biodiversity, we refer to [4].

The expert group [2] described a law for species abundance data: “When plotted as a histogram of number (or percent) of species on the y-axis vs. abundance on an arithmetic x-axis, the classic hyperbolic, ‘lazy J-curve’ or ‘hollow curve’ is produced, indicating a few very abundant species and many rare species (Fig. 1a). In this form, the law appears to be universal; we know of no multispecies community, ranging from the marine benthos to the Amazonian rainforest, that violates it” (in our Supplement A, we make some remarks to the meaning of ‘community’ and ‘sample’). Several terms are used for a ‘hollow curve’: the distribution is (right)-skewed, long-tailed, has extreme values, shows rare events. Is there a simple meaningful equation for ‘the’ species abundance distribution? One is inclined to think so if there is a ‘universal law’. Review [2] is steeped with the idea of a relatively simple equation for SADs, but it presents the opposite too, that different communities have rather different SADs and that groups of species within a community have different SADs, making the community’s SAD a mixed one [5]. The SAD is mostly treated as a histogram, based on the binning of data into frequency classes (for a probability mass or density function). However, the SAD can be illustrated as a rank abundance or Whittaker plot (see [2], their Fig. 1c; see also [6] [7]). Ranked data are used for exploratory data analysis. Rank-size plots and (cumulative) probability plots are strongly related [8] [9] (and see our Supplement B). SADs bear similarity to distributions in other fields of science [10] [11] [12].

Long-tailed distributions of natural and manmade phenomena, in rank-size form (where ‘size’ can be read as ‘abundance’), often show power law behaviour [13] (and see Wikipedia headword ‘Power law’). The ubiquitous power law has been considered for
species abundance data. However, ideal power law behaviour is absent or rare: data points do not lie in a straight line in a log/log plot \[14\] \[15\] \[16\]. For this reason, the interest in the power law for species abundance data seems to have vanished. However, the imperfect power law behaviour in other fields of science is well documented \[8\] \[17\]. Paper \[17\] is of particular interest. It revisits the data and analysis in a seminal paper \[13\] and concludes that 9 of 24 data sets conjectured to follow a power law actually do not.
**Figure Legend**

**Rank abundance plots for 12 data sets.** For each set, three samples at different scales of integration are shown: an average primary sample, an average composite sample of a subset, and the composite sample of the whole set (full-resolution file available as raw data). Data points in the log/log plots do not form straight lines; thus, ideal power law behaviour is absent. Stretched exponentials were fitted (drawn lines).

**Results & Discussion**

Rank abundance plots were made, shown in Fig. 1. Data points in the log/log plots do not form straight lines, ideal power law behaviour is absent. The data points indicate curved lines, concave in almost all cases. As an alternative to the power law, the stretched exponential function was fitted. The parameter values of the fitted function, for the composite samples of the complete sets only, are given in our Supplemental Table B. No advanced statistics was applied. No comparison with other functions/models was made. The actual data show some deviations from the fitted curves, but the overall result is visually satisfying.

For further work, advanced statistics are required; we refer to [17] [18]. For model comparison and selection, we refer to [19] [18] [20]. For allies of the power law, we refer to [8] [21] [22] (especially chapter 4) [17] [23] [24] [25].

A challenge lies in a remark in the review of the expert group [2]: "Starting in the 1970s and running unabated to the present day, mechanistic models (models attempting to explain the causes of the hollow curve SAD) and alternative interpretations and extensions of prior theories have proliferated to an extraordinary degree". The power law and its allies are often considered for the degree or connectivity distribution of networks. Species abundance data are retrieved from ecological communities that are networks [26]. However, a network topology behind species abundance data is not immediately clear. Species abundance data are reminiscent to data of food webs. For instance, the interactions between fruits and frugivorous birds can be presented in a cross-table of fruit species × bird species (data of [27] in [28]). From such a table, one can summarize the number of connections for the fruiting species with the bird species and vice versa: two connectivity distributions. The one dealing with the number of interactions for fruiting species over bird species is reminiscent of an assemblage of fruiting species, ‘sampled’ by birds. Networks can be generated by a process called preferential attachment (assortative mixing, assortativity [13] [29] (his Fig. 1, as well as a video in its Supplement S3), and see also [30] and Wikipedia headword 'Preferential attachment'). We suggest to link the quest for a simple distribution equation for the SAD with network research [31] [32].

Not only ‘our’ data but, quite often, other data as well consist of sample × species crosstable. Such tables provide for the opportunity to merge samples into a composite sample for a subset, or the whole set, as we did. Another opportunity is to study ‘within species over samples’ characteristics. We point to the abundance-occupancy relation [33], to Taylor’s law (fluctuation scaling [34] [35]) and to sampling theory [36]. Hopefully, all patterns can be integrated and applied for analysis with resampling statistics [37] (see also Wikipedia headwords ‘Nonparametric statistics’ and ‘Resampling (statistics)’) to obtain robust results, especially on the SAD.

Some of ‘our’ data sets provide for a time or a spatial series (fine scale: Mushrooms, Fish, Crustaceans, Fish+Crustaceans, Trees and Rodents). This makes them eligible to study the process of accumulation (‘sample’ growth, collectors curve, and species area relation [38] (see also [32]) and to look for autocorrelation). Seasonality aspects have already been described for the Mushrooms set [39], for Fish [40], and for Rodents and Annual plants [41] [42] [43]. Spatial patterns for the Trees set have been described in [44]. Our result makes a robust impression (12 systems at three scales of integration) that the stretched exponential [8] [17] is a possible alternative for the power law. This result may stimulate others to pick up again the power law, its allies, and their generative properties for species abundance data.

Our statistics are traditional and limited in scope. Neither advanced goodness-of-fit testing is done nor statistical comparison with other models is made. We did not study, just considered, generative processes.
Ideal power law behaviour is absent in the data sets. Data points in log/log plots show curvature, concave in almost all cases. Fitted stretched exponentials meet this curvature. Advanced goodness-of-fit testing, model comparison/selection, and generative processes need to be done, expertise that we do not master and for which we seek collaboration.

Additional Information

Methods and Supplementary Material
Please see https://scienccematters.io/articles/201612000009.

Ethics Statement
Not applicable.

Citations


